



## Phonology constrains the internal orthographic representation

LEONARD KATZ & STEPHEN J. FROST

*Haskins Laboratories, New Haven, USA and University of Connecticut, Storrs, USA*

**Abstract.** Four experiments explored the composition and stability of internal orthographic representations of printed words. In three experiments, subjects were presented on successive occasions with words that were consistently spelled correctly or were consistently misspelled. On the second presentation, subjects were more likely to judge both kinds of words as correctly spelled than on the first presentation, suggesting that their preexperimental orthographic representations had been altered to match what they had seen on the first presentation. However, only misspellings that were consistent with the correct phonology were accepted; spellings that altered the phonology were rarely accepted, suggesting that some parts of the orthographic representation are less stable than others. Also, subjects' reliance on orthographic vs. phonological memory when judging a word's spelling was affected by the kinds of other misspellings in the list. Lists that contained some phonologically implausible spellings for real words (e.g., \**assostance*) induced subjects to rely more on phonological plausibility when judging the correctness of other words in the list and less on orthographic memory. An individual grapheme in an internal orthographic representation was unstable when there were many phonologically acceptable alternatives for it. The results are contrary to the view that the strength of an internal representation is uniform across all its graphemes and is a function only of visual experience with the printed form. Results were interpreted in the context of a theory that considers spelling knowledge to be a by-product of the reading process, a process that involves phonological analysis.

**Keywords:** Orthography, Spelling, Reading, Word recognition

### Introduction

Most people who are literate in English have less than perfect spelling. The problem arises because the English writing system does not represent the pronunciation of a word in a simple and consistent manner and, indeed, was not meant to do so. Instead, the writing system – the orthography – has a major focus on the morphological information carried in the word. Occasions when the orthography reflects the consistency of meaning rather than the phonology can be seen as an accommodation to the complex phonology that exists in English (Katz & Feldman 1996; Katz & Frost 1992; Mattingly 1992). The pronunciation of a root morpheme or inflection often changes as a function of the syntactic or phonotactic context (as in *magic*, *magician*). Languages tend to get the orthography they deserve, which is to say that

there is usually a rational match between a writing system that evolves and the spoken language it represents (Katz & Frost 1992). Many of the world's written languages have less complicated orthographies than does English because they have a less complicated phonology. For many languages (e.g., Turkish, Finnish, Serbo-Croatian), to know how to pronounce a word is to know how to spell it. For the writer of English, in contrast, word-specific information about a spelling is required.

How can we characterize the cognitive (i.e., internal) representation of a word's spelling? Two kinds of information have been suggested as being available to the speller: (1) associations between phonemes or phoneme clusters and letters, and (2) associations between spoken morphemes (including words) and whole printed morphemes. Treiman (1992), in discussing the process of associating phonemes to letters, argued that children often use graphemic memories of clusters, units larger than a phoneme but smaller than a closed syllable, in spelling. These include intrasyllabic units such as onset clusters (e.g., *bl-* in the word *blow*) and rimes (e.g., *-ight* in the word *fight*). If the speller is able to parse the spoken word phonologically, he/she may be able to associate each component of the parsing with a corresponding spelling. But often the relation between phonology and orthography is one-to-many. For example, given the spoken word *fight*, there are two ways to spell the rime, *\*fite* and *fight*, but only one of them is orthographically conventional. Thus, the speller must remember orthographic information specific to the given word (Bruck 1993; Brown 1988) although a speller may use both this kind of information and information of phoneme-grapheme correspondences (Kreiner 1992). Consistent with the idea that there is word-specific information, Rapp, Folk, Botelet and Skultety (1998) offered evidence that the internal orthographic representation consists of two kinds of information: structure and texture. Structure consists of organizational information beyond the level of the identity of the graphemes and their order. This includes information that marks a letter as a doubled letter and possibly information concerning the grouping of letters into syllable or syllable-like units. Texture refers to the memory strength with which the specific graphemes are represented. Grapheme signal strengths may differ among themselves within the same word.

Additionally, some spelling knowledge must consist of learned general regularities such as the orthotactic characteristic of English (e.g., the letter cluster *ck* can occur at the end of a morpheme – *duck* – but not at the beginning – *\*ckup*). Morphological knowledge can also be used to spell, e.g., the knowledge that plurals are formed by suffixing the letters *-s* or *-es* or the convention that interrogative words may begin with a *wh-*. But much spelling knowledge must be learned by rote, either because the spelling is idiosyncratic or, if

there are historical or linguistic principles behind the spelling, the principles are unknown to the learner. For example, the words *signature* and *sign* are morphologically related; knowing how to spell the former can help one to spell the latter – but few spellers will have knowledge of the relationship. For most learners, the spelling of *sign* will be learned by rote.

The experiments presented here address the general question of the nature of the cognitive spelling representation and its acquisition. The main task is to account for why all letters in a word are not equally likely to be misspelled. For example, given the spoken word *parable* few if any literate people will misspell the initial (consonantal) phoneme in each syllable but spelling errors may occur on the vowel, particularly on the unstressed vowel. The problematic letters are not, of course, always the vowels; consonants may also be vulnerable, as in the consonant gemination in the word *misspelled*; many people will accept \**misspelled* as correct. Why are some letters problematic and not others?

It has been proposed that those letters whose relation to the word's phonology are most ambivalent (or multivalent) will have the weakest orthographic representation (Kreiner & Gough 1990). *Misspelled*, whether written with a single or double-*s* will be phonologically acceptable, given the indefiniteness of English syllable boundaries (and ignoring morphological considerations). Vowel phonemes in unstressed syllables (as the schwa in *acceptable*, *possible*, and *author*) are vulnerable because they have a particularly ambiguous relation to the printed vowel. The letter *i* in *possible* is only one of three vowels that could have been used to produce the correct pronunciation (the others are *a*, *e* and *u*).

How is the spelling memory of *misspelled* formed? We propose that phonology is involved in forming the internal orthographic representation via the orthography-phonology relationship; we propose that the process of forming an orthographic representation is not completely based on perceiving, coding, and storing visual orthographic information alone. The claim is that the internal orthographic representation is not formed simply as a passive reflection of the visuo-spatial characteristics of the print but, rather, the reader's knowledge of the relations between orthography and phonology shapes the internal representation. Thus, some of the word's letters are more likely to be encoded and retained as graphemes in memory than others. The letters that will be favored in the internal representation are those that have the least ambiguous mapping from phonology to print.

The research reported here probed the structure of cognitive spelling representations by observing the kinds of misspellings that are accepted as correct by subjects – and the kinds of misspellings that are never accepted as correct. The logic of this approach is that those misspellings which, in

spite of being incorrect, are accepted by a subject as correct reveal weaknesses or instabilities in the subject's internal orthographic representations. Conversely, misspellings that are never accepted reveal structural strengths of the representation.

Consider a word that is presented twice to a subject for a spelling evaluation. Some misspelled words will have the following pattern: The word will be (correctly) judged as a misspelling on its first presentation but (incorrectly) judged to be a good spelling on its second presentation. For these words, only one exposure to a misspelling will be sufficient to make the misspelling subsequently 'look good' to the subject, i.e., match his/her internal orthographic representation. The subject's preexperimental orthographic representation of that word may have been deficient in some way; either the subject had no memory at all of the spelling on first exposure and was guessing randomly or the subject did have a preexperimental orthographic memory which was altered by the misspelling observed on the first presentation.

The kinds of cognitive representations that will be vulnerable to alterations of this kind will contain a weakly encoded grapheme in the same letter location as the misspelling. Thus, the kinds of misspellings that are at first unacceptable but then become acceptable inform us about which letters were encoded weakly (if at all) in the subject's preexperimental orthographic representation. The same is true of misspellings that are accepted by the subject on first presentation. The subject's mistakes tell us which letters in the spelling are vulnerable to errors. By the terms vulnerability or instability we mean that the mental orthographic representation is initially ill-formed or weakly formed. As we noted above, vulnerability or instability occur in a systematic, principled, manner; the most vulnerable spellings contain a letter that is only one among several likely letters that can express the phoneme. One consequence of this is that the precise encoding of the grapheme is not critical for the reader's knowledge of the word's phonology because the letter's redundancy is high or, inversely, its information value is low.

*Does reading determine spelling knowledge?* As the previous discussion suggests, we propose that spelling knowledge is mainly acquired during the reading process, as a by-product of that process. In order to explain our theoretical approach to spelling, we must, therefore, first sketch our approach to printed word recognition. There are three contrasting points of view that characterize the process of how printed words are recognized: visual, phonological, or a mixture of the two. This issue has a long history in cognitive psychology (see, for a discussion, Frost 1998).

All three approaches assume that the task for the beginning English reader is to master what Gough, Juel and Griffith (1992) refer to as 'the cipher' between the printed and spoken word, the set of correspondences between letters (or letter clusters) and phonemes (or phoneme clusters). That the cipher must include letter clusters seems clear from the fact that we can learn to make different pronunciations to the same letters when they are in different contexts as, for example, the different initial phonemes in the words *choose* and *chord* even though both words begin with the same three letters. Thus, the cipher eventually learned by all skilled readers is a set of multilevel transforms that carry the printed form into the spoken. In addition to this cipher, we will include the converse mapping that carries spoken forms into their spellings (top-down).

The three points of view presented below differ in the degree of analysis of the printed word that is required of a skilled reader in order to access his/her phonological representation of the word. For the first view presented, the reader transforms letters and letter clusters within the printed word into phonemes which are then combined to form a phonological probe that addresses (and activates) the internal phonological representation (the mental lexicon's knowledge of the spoken form). In this view, the reader is a 'decoder', if we expand the notion of decoding to include rapid fluent analysis performed with little attention. Purely orthographic information, such as information about the actual spelling of a word, is said to be utilized only in the later stages of word recognition (e.g., as a spelling check), after candidate phonological entries have already been activated in the lexicon. Arguments for this position, based on tests of normal skilled readers, have been offered by Lukatela and his associates (Lukatela et al. 1989, 1994a, b), Perfetti and his associates (Perfetti, Zhang & Berent 1992), Stone and Van Orden (1994) and others.

In contrast to the first view which requires the reader to analyze the printed word into its component letters or letter clusters, the second approach assumes that the orthographic representation is not analyzed into phonological components but, instead, makes contact as a whole with the phonological lexicon. Caramazza, Berndt and Basili (1983) and his colleagues (e.g., Miceli, Benvegnu, Capasso & Caramazza 1997) argue for this second (nonanalytic) view; they have presented evidence from patients with various kinds of neurological damage, that printed word recognition does not obligatorily depend on phonological support and, in fact, recognition may be a process that is exclusively based on whole word orthography. For Caramazza, the skilled experienced reader does not need phonology, at least for words that he or she has experienced before. Support for this position also comes from Hanley and McDonnell (1997) who present data from an aphasic patient whose phonological representations have been apparently severely

compromised but who, nevertheless, reads and comprehends (and spells) normally.

The third view is a hybrid of the first two; familiar words are recognized orthographically, i.e., either by the whole orthographic form or by a combination of letter clusters. Less familiar words, however, must be analyzed first, at a smaller grain size, into phonological information which becomes a probe used to access the phonological lexicon. Evidence for two such active modes of word recognition comes in part from comparisons across orthographies that differ in the complexity of the cipher (e.g., Frost, Katz & Bentin 1987; Oney, Peter & Katz 1997) and from experiments in which word naming (reading aloud) is observed to be affected by phonological variables (e.g., Rastle & Coltheart 1998). For example, Katz and Feldman (1983) presented data suggesting that readers of Serbo-Croatian, an orthography that has a simple and consistent relationship between letters and phonemes, were more likely to recognize printed words via the phonology than were readers of English. Because the simpler Serbo-Croatian orthography has no ambivalence in its connections between letter and sound, activation of phonology from print can be faster than in English where ambivalent connections prevail.

*The representation of spelling knowledge.* It will be heuristically useful to consider the two extreme views, i.e., the phonologically analytic and the nonanalytic. If, as the analytic view has it, phonology is both obligatory and exclusive, we might expect that readers would not devote cognitive resources in order to remember the orthographic information in its full detail. Because the phonological form is necessary and sufficient (in this view) to address meaning, then the orthographic information that is needed to get to the phonological representation in the lexicon can be discarded once the phonology has been activated. Thus, spelling becomes less relevant to the word recognition. However, it can not be the case that orthography is completely irrelevant, because we should not be able to remember that the word *fish* is not spelled \**phish* or that *eat* is not spelled \**eet*. Therefore, at least some information must be retained as a graphemic memory. However, for addressed phonology theorists (e.g., Lukatela et al. 1994b), internal graphemic information is activated only after lexical access has occurred via phonological analysis of the print.

As we said earlier, spelling information is obviously valuable because it carries morphological meaning and, therefore, can disambiguate homophones; it is useful to know that we are reading the word *heal* and not word *heel*. Thus, the extreme form of this approach, in which orthography is completely irrelevant to lexical lookup, must be wrong. But, as we stated above, even proponents of the analytic view allow for postlexical utilization

of orthographic information as pure orthography, in order to account for our ability to discriminate homophones like *heal* and *heel*. The less extreme version of this approach is viable; Frost (1998) gives an account of the nature of orthographic representations under less extreme assumptions.

Consider the other extreme. If access to meaning involves orthography exclusively – if phonology plays no role – then the reader must retain as much of the orthographic information as he or she can precisely because it is this information that discriminates among meanings. An important prediction of such a model is that a reader's failure to completely encode all the letters of a word in orthographic memory (i.e., spelling errors) should be neutral with respect to the phonology. That is, the letters that fail to be encoded in orthographic memory and the letters that are encoded should be predictable from visual-perceptual laws but not from phonological-perceptual laws. In this scenario, the letters that tend to be misspelled might be letters that are laterally masked by other letters (i.e., internal letters) but not the initial and final letters of a word. However, the internal letters should not be further distinguished among themselves. Importantly, misspelled letters should not be predictable from considerations of the grapheme-phoneme cipher. In misremembering the spelling of the word *misspell* it is as likely for the letter *p* to be forgotten as the letter *s* that precedes it. But it is obvious from the observation of spelling errors that the preceding conjecture is false. Errors are not aphonological but, rather, are nearly always errors that produce a homophone of the correct spelling. To use Bosman and Van Orden's (1997) example, most Americans will misspell the name of 'the Indian nationalist and spiritual leader who developed the practice of nonviolent resistance ... known as Mahatma \_\_\_\_'. This error occurs in spite of a reasonable frequency of exposure to the name. Apparently, the error occurs because the word's pronunciation is unaffected by it. [The correct spelling is Mahatma Gandhi.]

From these considerations, it seems unlikely that either the orthographic or phonological process<sup>9</sup> is exclusively employed in printed word recognition. We can speculate that some of the orthographic information (but not all) is processed into a graphemic representation, thereby forming a partial graphemic memory of the printed word. In a slightly different version of this conjecture, all the orthographic information, rather than only a part of it, is processed initially but not all letters in a word have equally stable encodings. Either way, however, the interesting question is whether the orthographic information that fails to be remembered is or is not predictable by the cipher, i.e., by the learned relationship that exists between graphemes and phonemes. If it is, then phonology can be said to play a role in determining which orthographic information fails to be stored and/or remembered.

*The relation between spelling and word recognition.* In discussing the question of the relationship between orthographic and phonological representations, it is necessary to consider a related question, the question of whether the act of spelling a word draws on the same information as the act of reading (recognizing) that word. (This is a valid question whether the recognition process involves phonology or not.) Bradley and Bryant (1979) maintain that reading and spelling are separate at least for beginning readers and for impaired readers but take no stand on whether this is the case for skilled readers. Hanley and McDonnell (1997) imply that both recognition and spelling need depend only on an orthographic representation. We suggest that reading and spelling both depend on graphemic *and* phonological information, although spelling a word is not simply the reverse of recognizing a word.

Observation and intuition suggest that spelling initially involves the execution of general sound-to-spelling knowledge (one side of the cipher). Such knowledge is sufficient to prevent misspellings such as \**ckat* for *cat* because the initial phoneme in the word is never spelled with the letters *ck* when it is in initial position although the phoneme is, in fact, spelled that way when it is in the final position in a word (e.g., the word *lack*). After a spelling has been produced, it is often examined by the speller, visually or in visual imagination. Reading (recognizing) the spelling informs the speller if it is consistent with his or her orthographic memory for the word and whether a change is required.

If the orthographic representation is unstable, what determines which part of the orthographic information is encoded successfully and which part is less favored? We suggested above that the word's phonology may affect this process (see, for an earlier, similar proposal, Kaminska & Dixon 1994). It is the linkage between print and speech in the orthography that is crucial. The idea can be refined as follows: the encoding of graphemes during the process of printed word identification is a function of the simplicity of the grapheme-phoneme relation; the simpler the relation, the more likely a stable graphemic encoding will result and, therefore, a stable memory. For example, consider the letter *b*. That letter is almost always given the same pronunciation, /b/, whatever the orthotactic context. There are only a few exceptions (e.g., *tomb*, *bomb*) where the letter has a different pronunciation value. (In the case of *tomb* and *bomb*, the alternative pronunciation is silence.) Moreover, the inverse relation is also simple; the phoneme /b/ is always written with the letter *b*.

The conjecture is that such letters require less processing than others, because of their relational simplicity. Processing is less costly because fewer alternative encodings need be evaluated by the system before a decision is



made about the phonemic value of the letter. In the inverse decision, fewer grapheme alternatives need to be considered when a spelling for the phoneme /b/ must be produced. In contrast, many other letters and phonemes have more complex linkages. Letters like the vowels in unstressed syllables, which all tend to be pronounced as schwa in English, or geminate letters, which are sometimes irrelevant to the pronunciation, will be the least favored for graphemic encoding. For example, least favored for successful stable graphemic encoding will be letters like the second a in *acceptable* or the geminate s in *misspell*. Of course, over time, with many exposures to the same word, even less favored letters in the spelling may be learned well. But even then, evidence should exist of greater stability for the simpler, more favored letters.

### Experiment 1

We attempted to demonstrate that the letters that are most vulnerable in a word's orthographic representation are those whose relation to the word's phonology are most ambiguous. The most ambiguous should be geminates and letters representing unstressed vowels. The word *affable* contains examples of both these: an internal double-f and an unstressed vowel in its second syllable. We observed instability in words like *affable* by looking for changes in a subject's evaluation of a given spelling over two successive presentations. If an internal orthographic representation is, in fact, unstable it might be possible to alter that representation with a single exposure to an alternative representation. Thus, spellings that looked incorrect the first time might look correct the second.

Each subject was presented with a set of words, some of which had been misspelled at one letter location. After a short delay, the same list of words was presented a second time. Of particular interest were response sequences in which a subject judged a word correctly on the first presentation but incorrectly on the second. For example, consider a misspelled word (e.g., \**mispell*) that may have been judged, properly, as misspelled the first time the subject saw it but on its second exposure, was judged as correct. The incorrect component in the initially presented spelling may have been substituted for the correct one in the subject's internal orthographic representation – if the subject's representation for that letter had originally been ill-formed or unstable.

Because we wished to measure response times for spelling judgments, we decided to use computer presentation of stimuli. However, we also wanted to validate this experimental technique against a nonspeeded spelling recognition test and against the traditional method of assessing spelling, i.e., spelling from dictation. Therefore, a second purpose of Experiment 1 was to compare

the three methods of assessing spelling performance: by computer, with reaction times as well as number of correct judgments, by a nonspeeded spelling judgment test, and by a test of spelling from dictation, the traditional method. The same subjects were given the tests of spelling from dictation and nonspeeded spelling judgment. This allowed us to assess whether the spelling judgment task preserved the individual differences that would be found in the traditional test of spelling ability, spelling from dictation.

### *Method*

*Stimuli.* Forty-eight high frequency words and 48 low frequency words were selected. Within each set, 24 words were altered so that they were misspelled and 24 were left correctly spelled. The mean Kucera-Francis frequencies for the four subsets were: low frequency correct, 6.45, low frequency incorrect, 6.70, high frequency correct, 101.16, high frequency incorrect, 96.83 (Kucera & Francis 1967). Appendix A contains two lists: the first presents the 24 low frequency correctly spelled and misspelled words (along with their correct spelling.) The second list contains the same information for the high frequency words.

*Subjects.* Students at the University of Connecticut participated in the experiment for course credit. There were 47 subjects in the paper and pencil group and 44 different subjects in the computer presentation group.

*Procedure: Paper and pencil presentation.* Subjects in the paper-and-pencil evaluation group were tested in 2 subgroups of 22 and 25 subjects. In the Dictation phase, subjects were given sheets of paper with numbered blank spaces written in columns. Subjects were asked to listen to the experimenter as she said each word (it was spoken twice), and to print the word, correctly spelled, in the blank space. At the end of the list, the papers with the subjects' spellings were collected and there was a 5-minute pause.

Then, unexpectedly, the Recognition phase began. Each subject was given a list of the 96 words printed (in the same order as had been dictated before). Half of the words were correctly spelled, half misspelled. The subject was told that some of the words on the printed list may have been spelled incorrectly; their task was to look at the word as spelled on the paper and judge their degree of confidence about the correctness of the word's spelling. Decisions were made on a 4-point scale printed next to each word: 1 – the word is spelled incorrectly, 2 – I think the word is spelled incorrectly but I'm not sure, 3 – I think the word is spelled correctly but I'm not sure, and 4 – the word is spelled correctly. There was also a category labelled 'Don't know', for use when they did not know the word.

*Procedure: Computer presentation.* Subjects in the computer group were presented, on each of 96 trials, with a 500 ms fixation point (an asterisk) followed by a 50 ms blank screen and then a word in the center of the screen. The word remained on the screen until the subject made the decision that the word was spelled correctly or spelled incorrectly by pressing one of two telegraph keys. If no response was made by 2500 ms, the stimulus terminated. There was a 1200 ms intertrial interval. The key labeled 'spelled correctly' was always under the subject's dominant hand. Subjects were told that speed of response was important even though accuracy should be maintained (Kreiner 1992, used a different kind of speeded assessment of spelling knowledge: a probe letter paradigm, similar to the classic probe-digit paradigm of Sternberg 1969). After 20 practice trials, subjects were given the opportunity to ask questions about the procedure. Then the sequence of 96 words was run without pause (Pass 1). Each subject was presented with a different random sequence of words. After the list was finished, there was a 5 minute pause and then, unexpectedly, the 96 trials were repeated, in the same order, for a second pass. Following the experiment, subjects were debriefed and were presented with a list of the experimental stimuli spelled correctly.

### *Results and discussion*

For the paper and pencil group, means were obtained for (1) the mean number of correct spellings from dictation, and (2) the mean recognition confidence. The mean response for words in each combination of Frequency (low/high) and Spelling (correct/misspelled) was calculated. For the computer presentation group, both mean reaction time (RT) for correct decisions and mean number of correct responses was calculated within each combination of Frequency and Spelling. For analyses based on subjects, means were calculated for each subject aggregated over the items to which the subject responded. For items analyses, means were calculated for each stimulus aggregated over subjects responding to that stimulus.

*Spelling from dictation.* The list was difficult enough to produce sufficient variability of individual differences for study. The average percentage of correct spellings for the dictation list was 72.8%, SD = 22.7%. There were several words that fewer than half of our subjects got correct. These words could be grouped into two broad groups: words that elicited errors involving (1) gemination and/or an unstressed vowel, i.e., errors related to phonology, and (2) words that elicited errors involving rote orthography, i.e., errors specific to orthography. Words of the first type were: *assessors*, *negligible*, *personnel*, *barricade*, *predecessor*, and *temperament*. Words of the second type were: *nuisance*, *camouflage*, and *truly* (the misspelling was *\*truely*).

Thus, there seems to be ample evidence that the proportion of spelling errors involving phonology is substantial (if not predominant) for the subjects we observed. In this regard, the college students we studied are similar to primary school children (Treiman 1993) and high school students (Shankweiler, Lundquist, Dreyer & Dickinson 1996).

*Correlations among the methods.* First, we correlated subjects' percentage correct on the two paper-and-pencil tests. A measure of the validity of the unspeeded spelling evaluation task (unspeeded recognition) is given by its correlation with the dictation task when subjects are the unit of analysis. The correlation was satisfactory,  $r = 0.58$ ,  $p < 0.0001$ . For low frequency words alone on the two tasks, subjects' scores correlated well,  $r = 0.70$ ,  $p < 0.0001$  and for high frequency words,  $r = 0.38$ ,  $p < 0.009$ . Thus, the subjects who took the two paper-and-pencil tests tended to perform similarly on spelling from dictation and unspeeded recognition.

It would be expected that reaction times in the computer task (speeded recognition) should be fast to words that were easy to spell correctly (in dictation) and these same words should also be given high confidence ratings in unspeeded recognition as having been spelled correctly. The stimulus, therefore, was the unit of analysis in our main assessment. Table 1 presents the correlations among the three methods. Note that Table 1 presents correlations for Pass 1 and Pass 2 of the computer test separately. All correlations are of at least moderate size and are statistically significant (all  $p < 0.01$ ), indicating that there is good agreement among the computer measures (RT and number correct), the unspeeded recognition test, and the number correct when spelling from dictation. These intercorrelations make it clear that all methods give similar results, whether subjects are asked to spell a word from dictation or to judge the acceptability of a printed spelling (whether in either a speeded or unspeeded decision). Words that are difficult to spell are also difficult to recognize as being correctly spelled.

#### *The effect of repetition: Computer procedure*

*Number correct.* For the computer presentation procedure, ANOVAs were performed on the number of correct responses. The factors were Frequency (low/high), Spelling (correct/misspelled), and Pass (one/two). The three-way interaction was significant, both by subjects ( $F(1) = 28.99$ ,  $MSe = 2.415$ ,  $p < 0.0001$ ;  $F(2) = 16.90$ ,  $MSe = 7.60$ ,  $p < 0.0001$ ). Figure 1 shows that for high frequency words there was little increase from Pass 1 to Pass 2 in correctly identifying a stimulus as misspelled or correctly spelled. For low frequency words, correct responses to correctly spelled words increased from Pass 1 to Pass 2. More interestingly, correct

*Table 1.* Correlations among three methods of assessing spelling: Spelling Acceptability Rating, Spelling from Dictation, and Speeded Decision (reaction time and number of correct responses on passes 1 and 2)

	Rating	Dictation	RT1	RT2	Ncor1
Rating	—				
Dictation	0.736	—			
RT1	-0.616	-0.435	—		
RT2	-0.714	-0.491	0.797	—	
Ncor1	0.813	0.531	-0.732	-0.797	—
Ncor2	0.847	0.579	-0.663	-0.808	0.920

responses to misspelled words decreased. That is, for low frequency words, a single exposure to a stimulus in Pass 1 was enough to make that spelling more acceptable on Pass 2. This result is similar to results found by Jacoby and Hollingshead (1990) and Kaminska and Dixon (1994), using different paradigms from the present one. In both of those studies, as in the present Experiment 1, exposing a subject to a misspelling affected subsequent spelling accuracy.

Thus, subjects' preexperimental orthographical representations for low frequency words were unstable or weak enough so that even a single stimulus presentation in Pass 1 could significantly alter the memory representation (whether the stimulus was correctly spelled or not). The mental orthographic representation of an item tended to change so as to match the most recent exposure.

*Reaction time.* Reaction times on correct responses were also examined. Figure 2 presents the main outcome. Inspection of Figure 2 indicates that responses were faster to correctly spelled words and to high frequency words. Furthermore, the frequency effect appears to be larger for correctly spelled words. Responses on Pass 2 were faster than on Pass 1 and reaction times for correctly spelled words became even faster from the first to second pass than misspelled words. An ANOVA supported these suggestions: the effects of Spelling and Frequency were significant by subjects and items analyses [ $F(1,42) = 96.32$ ,  $MSe = 52997.26$ ,  $p < 0.0001$ ;  $F(1,92) = 114.06$ ,  $MSe = 27044.66$ ,  $p < 0.0001$ ;  $F(1,42) = 60.18$ ,  $MSe = 13413.24$ ,  $p < 0.0001$ ;  $F(1,42) = 9.50$ ,  $MSe = 27044.66$ ,  $p < 0.003$ , respectively]. Their interaction was also significant [ $F(1,42) = 7.27$ ,  $MSe = 15075.51$ ,  $p < 0.001$ ;  $F(1,92) = 6.31$ ,  $MSe = 27044.66$ ,  $p < 0.02$ ]. Pass was significant [ $F(1,42) = 12.61$ ,  $MSe = 37655.23$ ,  $p < 0.001$ ;  $F(1,92) = 38.38$ ,  $MSe = 6923.74$ ,  $p < 0.0001$ ].

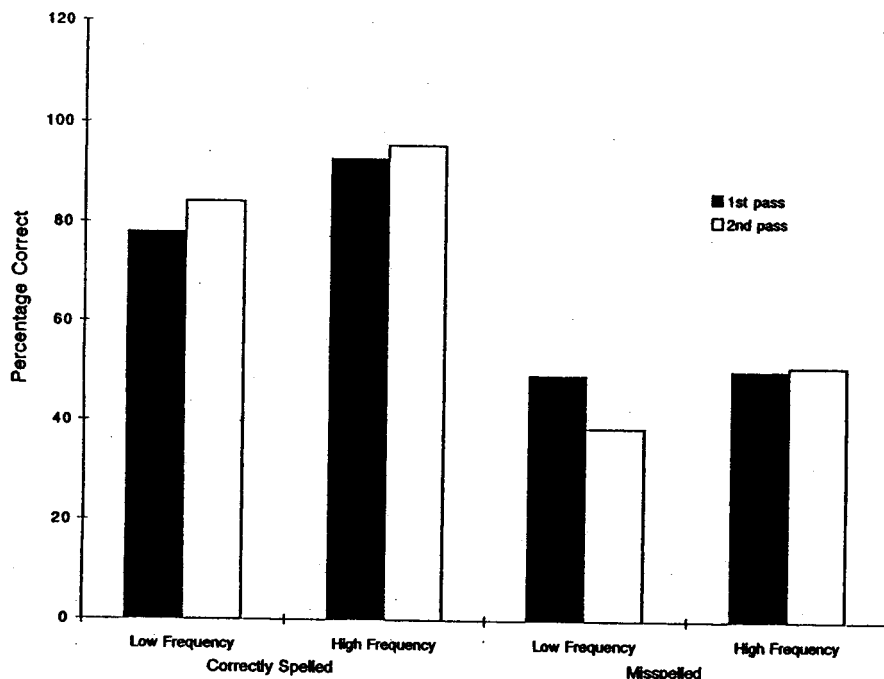


Figure 1. For computer presentation, the percentage of correct responses for low and high frequency words presented as correctly spelled or misspelled, on first and second passes.

Finally, the interaction Spelling  $\times$  Pass was significant [ $F(1,42) = 7.27$ ,  $MSe = 15075.50$ ;  $F(1,92) = 4.32$ ,  $MSe = 6923.74$ ,  $p < 0.05$ ].

The general reduction in judgment time for correct responses on the second pass indicates the advantage resulting from subjects becoming familiar with the stimuli on the first presentation. Of greater interest is the interaction of Spelling  $\times$  Pass. The improvement in speed of response (from Pass 1 to Pass 2) was less for misspelled words. The second time the misspelled word appeared, it may have looked better and, although the subject made a correct judgment based on his/her correctly spelled mental representation, that mental representation had been weakened by the previous exposure to the misspelled form. This result suggests that some mental orthographic representations had been altered by a single exposure to a misspelling.

## Experiment 2

Experiment 1 demonstrated that one exposure to a misspelled word can significantly increase acceptance of that misspelling on a second test. Further, Experiment 1 demonstrated the validity of the speeded recognition task. The

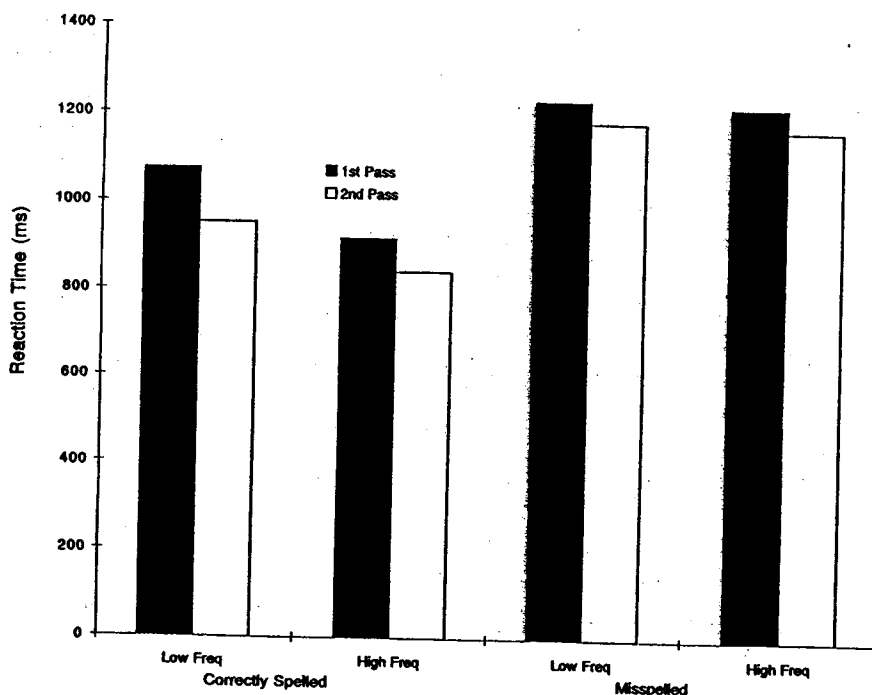


Figure 2. For computer presentation, reaction time on correct responses for low and high frequency words presented as correctly spelled or misspelled, on first and second passes.

latency of a spelling acceptance judgment can be a sensitive measure of the subject's confidence in that judgment.

Experiment 2 was designed to study the effect on the internal orthographic representation of additional exposures, beyond the first. Of main interest was how additional exposures continue to alter the acceptability of a spelling. Conventional learning theory predicts an effect that is cumulative, increasing, but with diminishing increments. In Experiment 2, each word, either consistently misspelled or spelled correctly, was repeated four times.

### Method

*Stimuli.* From the words in Experiment 1, a subset of 74 were selected; the words that were retained were those that had high error rates. Of these, 37 were spelled correctly and 37 were misspelled. There were no longer equal numbers of high and low frequency words. Instead, low frequency words predominated although there was still a wide frequency range. The Kucera-Francis ratings for correctly spelled words ranged from 1 to 392 (mean = 58.59, SD = 80.25) and for misspelled words the range was 1 to 222

(mean = 26.00, SD = 44.69). For the present experiment, it was judged more important to select stimuli which would produce errors (so that the effects of three successive presentations could be studied effectively) than to equate for frequency. Nevertheless, a posthoc analysis by frequency will be presented.

*Subjects.* There were 42 subjects; all participated as part of a course requirement at the University of Connecticut.

*Procedure.* The apparatus and procedure were similar to the computer procedure in Experiment 1 except that each stimulus was presented 4 times at random intervals continuously throughout the subject's single session – there was no separation of the first and subsequent presentations into distinct phases, as in Experiment 1. Thus, the 296 trials were run as one continuous series. The order of appearance of each stimulus was randomized across subjects. Not only was the trial number of the first presentation for a word random, all three subsequent exposures were also randomized for each subject. For a given word, the minimum lag length could be zero (immediate repetition, no intervening items) and the maximum could be 292 (when the word was presented on trial 1 and its second presentation occurred on trial 294). Aggregated over all the words presented to a subject, this resulted in a near normal distribution of distances between successive presentations for each subject. For each word, the actual trial number of its first pass and its three subsequent presentation trial numbers were recorded for each subject in order to be used in the analysis.

### *Results and discussion*

The number of trials intervening between successive presentations of a word was fully randomized. Both correct and misspelled words had, on average, nearly identical lag distributions. For correctly spelled words, the mean trial of first presentation was 58.20; for misspelled words, it was 59.61. The mean trial numbers for the second, third and fourth presentations were, for correct and misspelled words respectively, 119.15 and 118.69, 178.57 and 178.52, and 237.58 and 237.68. The main question we asked about the effect of repeated presentations is addressed in Figure 3, which presents the percentage of correct responses for correctly spelled and misspelled words over four passes.

Figure 3 shows that correctly spelled words were accepted 73% of the time on the first presentation and the percent acceptance increased further with subsequent presentations (83%, 86%, 87%) over the three successive exposures. For stimulus words that were misspelled, the percentage of correct response was 46% on the first presentation and decreased only slightly (43%,



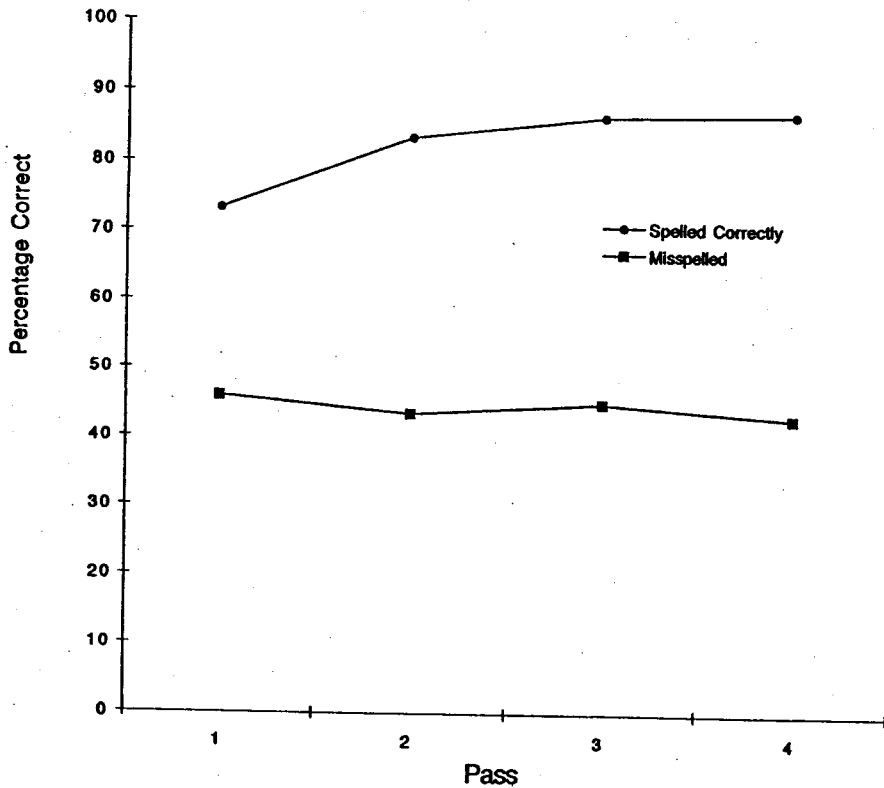


Figure 3. Percentage of correct responses for words presented as correctly spelled or misspelled over four successive passes.

45%, 43%) over the three successive presentations. ANOVA on subjects and items confirmed this observation; the interaction was significant [ $F(3,123) = 27.62$ ,  $MSe = 0.522$ ,  $p < 0.0001$ ;  $F(3,216) = 19.75$ ,  $MSe = 0.08$ ,  $p < 0.0001$ ].

Recall that stimuli were an uneven mixture of low and high frequency words. Nevertheless, we subjected the data to a posthoc analysis by frequency. We adopted a Kucera-Francis criterion of 30 and divided stimuli into words of low frequency (below 30) and high frequency. For misspelled stimuli there were 29 words that were low frequency and 8 that were high. For correctly spelled words, 20 were low frequency and 17 were high. For each of these four groups, *t*-tests were run to assess the significance of change in percent correct between Pass 1 and Pass 4, as a simple measure of change over time. For misspelled words, only the low frequency group showed a significant decrement in percent correct [ $t(28) = 2.11$ ,  $SE = 0.023$ ,  $p < 0.05$ ]. Over the four passes, percent correct responses were 48%, 43%, 45%, 43%. The high frequency group's mean showed the opposite pattern: it

actually increased between Pass 1 and Pass 4 (the four passes were 38%, 45%, 44%, 42%) but, perhaps because of the low power of this test, the difference between Pass 1 and Pass 4 was not statistically significant. In contrast, both high and low correctly spelled groups of words showed significant improvement in percent correct. For low frequency correctly spelled words the increase over the four passes was 67%, 80%, 83%, 85%. The difference between Pass 1 and Pass 4 was significant [ $t(19) = 6.80$ ,  $SE = 0.027$ ,  $p < 0.001$ ]. For high frequency correctly spelled words, the improvement was 81%, 88%, 90%, 89% [ $t(16) = 3.32$ ,  $SE = 0.025$ ,  $p < 0.01$ ]. Thus, the results of Experiment 2 are mostly in accord with those of Experiment 1 (i.e., on successive presentations, there is a tendency to adopt the spelling or misspelling of the previous presentation) with the exception of high frequency misspelled words. However, the data for high frequency misspelled words are problematic in the present experiment. Not only is the power of the statistical analysis low, but the particular high frequency words that were used may or may not be representative of such words in general. It remains for further research to address the issue more effectively than the present experiment.

Latencies of correct responses showed that responses were about 100 ms faster to correctly spelled stimuli than to misspelled stimuli and that latencies declined over presentations linearly from a mean of 1001 ms on the first presentation of a correctly spelled word to 800 ms on the fourth presentation. ANOVAs showed that there were strong main effects of Spelling and Presentation but no significant interaction [Spelling  $F(1,41) = 81.61$ ,  $MSe = 8026$ ,  $p < 0.0001$ ;  $F(1,72) = 71.42$ ,  $MSe = 60574$ ,  $p < 0.0001$ ; Presentation  $F(3,123) = 60.00$ ,  $MSe = 3780$ ,  $p < 0.0001$ ;  $F(3,216) = 23.21$ ,  $MSe = 16660$ ,  $p < 0.0001$ ].

Of major interest are the reaction times for incorrect responses, i.e., responses where subjects accepted a misspelled word or rejected a correctly spelled word. Figure 4 presents these latencies. The figure suggests that RT for accepting a misspelled word as correct was faster than rejecting a word that was correctly spelled. With successive presentations, latencies of incorrect acceptances decreased monotonically. However, latencies to incorrect rejections change monotonically over presentations: time to reject actually increased substantially from the first to the second presentation then decreased with further presentations. ANOVAs on reaction times supported these suggestions. The interaction of Spelling and Presentation was significant [ $F(3,123) = 4.69$ ,  $MSe = 44197$ ,  $p < 0.004$ ;  $F(3,168) = 3.69$ ,  $MSe = 39108$ ,  $p < 0.02$ ]. The main effects of Presentation and Spelling were also strongly significant.

The nonmonotonic pattern for incorrect rejection RTs over repetitions can be interpreted to indicate that a subject who thought that a correctly spelled

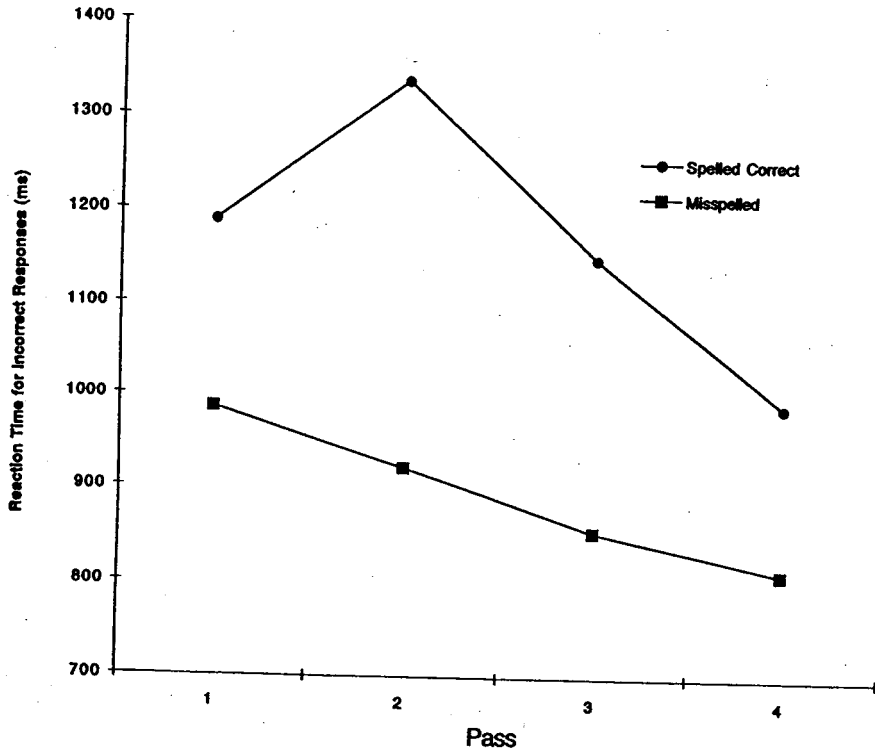


Figure 4. Reaction time on incorrect responses to words presented as correctly spelled or misspelled over four successive passes.

word had been spelled incorrectly on its first presentation was less certain when the exact same spelling reappeared for a second time. For these stimuli, it can be assumed that the subject had been exposed to the correct spelling preexperimentally and had a weak memory of that correct spelling. Although that memorial representation was not strong enough to prevent a decision that the word was misspelled, the additional experimental exposure to the correct spelling apparently increased the original representation's strength so that, on second presentation, the subject's second response was less certain and therefore slower than the first. However, on subsequent presentations, this inhibitory effect disappeared. Thus, the effect of exposure to the correct spelling had its strongest effect on the first exposure and, contrary to our expectation, there was no cumulative effect. The absence of a cumulative effect after the second exposure may show only that judgments are affected by the subject's memory of his/her most recent response. That is, the subject may be remembering his/her last response (that the word was misspelled) rather than evaluating the memorial representation. If so, then longer inter-

presentation lags might reveal a cumulative effect. Alternatively, it may be the case that spellings that are repeatedly judged to be misspelled are somehow blocked from memory storage. This would suggest that there is a mechanism to protect information that is repeatedly labeled as erroneous from being stored permanently.

A result obtained by Brown (1988) is partly consistent with the present findings. Brown presented college students with stimuli that were identified as misspelled and asked the subjects to rate each word on how closely it resembled the correct spelling (which was not present). Each list of misspelled words was presented three times. On the second presentation, a misspelled word was rated as more closely resembling the correct form than the rating on the first presentation. Brown's results differed from the present experiment in that ratings on his third presentation showed further acceptance of the misspellings. Both experiments agree, however, in demonstrating that only a single exposure to a misspelling may be sufficient to alter an orthographic memory.

### Experiment 3

In the first two experiments, the correct spelling was either present or absent on each trial. In the third experiment, we used a two-alternative forced-choice (2AFC) procedure in which, on each presentation, the correct spelling was paired with a misspelling and the subject's task was to choose the correct spelling as quickly as possible. The main purpose in evaluating a 2AFC procedure was to correct for a complication in interpreting the reaction time results of Experiments 1 and 2. The speed of a response to a stimulus word is interpreted as an index of the strength of the subject's internal spelling representation for the correct spelling: faster responses for stronger representations. However, when a large proportion of the subject's responses are in error, it becomes more probable that the average reaction time to correct responses contains some responses that were merely correct guesses. To that extent, the average reaction time is a less accurate index of the strength of the subject's orthographic representation. Thus, we sought to use a technique that might produce nearly perfectly correct responding.

The disadvantage of the 2AFC procedure is that the subject is presented with the correct spelling at the same time he/she is presented with a misspelling; this prevents the experimenter from differentially strengthening the misspelling as we did in the earlier two experiments. Instead, because both correct and misspelled forms of the word are reinforced, the effect on the subject's mental orthographic representation is unpredictable. Therefore, the 2AFC is not useful for observing the effects of exposing the

subject to a misspelling. Nevertheless, 2AFC should be useful for comparing the comparative effects of *different kinds* of misspellings. Each type of misspelling can be pitted against a common alternative, the correct spelling. The 2AFC paradigm should be a sensitive way to assess the composition of the mental representation by determining which type of misspelling has the greater 'attraction'.

In Experiment 3, two types of misspellings were studied: misspellings in which the changed letter rendered a different pronunciation to the stimulus (an unphonological misspelling, such as *\*sunitary*, paired with the correct spelling *sanitary*) and misspellings in which the changed letter did not change the pronunciation (a phonological misspelling, such as *\*sanitery*, paired with the correct spelling *sanitary*).

### *Method*

*Stimuli.* Two lists of 48 trials each were constructed. Each item consisted of a pair of words, on the same line. One of the pair was spelled correctly and the other was misspelled by changing a letter in either an unstressed or stressed syllable. In List A, half the items contained a changed letter in a *stressed* syllable, producing an unphonological misspelling, i.e., one which, if read aloud conventionally would produce an incorrect pronunciation. This misspelling was paired with the correct spelling, as in the pair: annihilation annihilation. The remaining items contained misspellings produced by changing a letter in an *unstressed* syllable producing phonologically acceptable misspellings (e.g., *\*admittance* – admittance). In List B, the same correct spellings appeared as in List A but the type of misspelling it appeared with was reversed so that each unphonological misspelling in List A became a phonological misspelling and vice versa (e.g., *\*anihilation* – annihilation and *\*admattance* – admittance). On half the phonological trials and on half of the unphonological trials, the misspelled word was presented on the left side; for the other 24 trials, the misspelled word was on the right side. The stimuli are presented in Appendix B.

*Subjects.* Subjects were 15 students at the University of Connecticut who were paid for their participation.

*Procedure.* Practice consisted of 10 trials and was followed after a short pause by the 48 trials of the main list. On each trial, a fixation point appeared for 500 ms, followed by the pair of spellings, side-by-side. The subject responded by pressing either the left or right telegraph key to indicate the side of the correctly spelled form. The trial was terminated when the subject responded

or when a 2 sec. timeout period was exceeded. The intertrial interval was 1 sec.

### *Results and discussion*

Mean reaction times to identify the correctly spelled form were 969 ms for unphonological trials and 1118 ms for phonological trials. The analysis of variance included Type of Misspelling as well as two methodological factors, List (A, B) and Side (Left, Right) of the correct spelling. Type of Misspelling was significant [ $F(1,13) = 25.38$ ,  $MSe = 29168$ ,  $p < 0.0002$ ], showing that the difference of 149 ms was quite reliable. No other terms were significant.

The error data paralleled the reaction time results; errors were smaller for unphonological misspellings (2%) than phonological (20%). This difference was significant in the analysis of variance [ $F(1,13) = 51.91$ ,  $MSe = 103.26$ ,  $p < 0.0001$ ]. the only other significant term was the interaction of Type and Side [ $F(1,13) = 5.22$ ,  $MSe = 3.13$ ,  $p < 0.04$ ]; phonological errors were slightly higher when the misspelling was on the left side.

Although Experiment 3 demonstrated the importance of grapheme ambiguity, it failed to provide an acceptably low error rate in all experimental conditions; only in the unphonological condition was the error rate acceptably low enough to consider reaction times unconfounded.

Clearly, subjects have more difficulty in determining the correct spelling when the alternative spelling is phonologically plausible. Alternatively, a misspelled letter in a stressed syllable is critical just because it disrupts the word's phonology; it is implausible. Why is a phonologically plausible misspelling sometimes acceptable? It must be the case that the correct grapheme is weakly represented in the word's mental orthographic representation (relative to the other graphemes in that word). The strength of a component grapheme in that orthographic representation may be inversely related to the grapheme's phonological ambiguity. That effect can occur, however, only if ambiguity in the letter's grapheme-to-phonology relationship plays a role in consolidating the orthographic representation at the time that a printed word is recognized. If so, then it is not the visual information alone from the printed letter that necessarily determines the orthographic representation but, rather, the phonological information in the reader's mental lexicon interacting with the printed letters. Together they produce an orthographic representation whose component graphemes are unequally consolidated. That is, phonology, in the phonology-to-spelling relationship, plays a role in acquisition and structuring of graphemes in an internal orthographic representation.